Ph.D. Title Proposal

on

AI based Phenotype and Crop Yield prediction in Climate Smart Agriculture Analytics

Submitted in partial fulfillment of the requirement for the award
of the Degree of

Doctor of Philosophy
in
Computer Engineering

by

Archana Kamlakant Salaskar

under the guidance of

Dr. Amiya Kumar Tripathy

Department of Computer Engineering
Mahavir Education Trust's
Shah and Anchor Kutchhi Engineering College Chembur,
Mumbai-400088
University of Mumbai December 2024

PROPOSAL FOR TOPIC APPROVAL TO BE SUBMITTED TO THE UNIVERSITY OF MUMBAI FOR THE DEGREE OF **PhD (SCIENCE AND TECHNOLOGY)** IN THE SUBJECT OF **COMPUTER ENGINEERING.**

Title of Topic	:	AI based Phenotype and Crop Yield prediction in Climate Smart Agriculture Analytics
Name of the Student	:	Archana Kamlakant Salaskar
Basic Qualification	:	M.E (Computer Engineering) B.Tech. (Computer Engineering)
In the Case of Ph.D. student Year of Passing of PET/NET/SET/GATE (Teaching Experience)	:	PET 2021
Name of the Research Supervisor/ Guide	:	Dr. Amiya Kumar Tripathy
Date of Superannuation / Retirement	:	07/10/2046
Name of the Research Center	:	Shah and Anchor Kutchhi Engineering College
Signature of Student	:	
Signatue of the Research Supervisor	:	

Seal of College

Head / Director / Principal
Research Centre

Index

Sr. no.	Content	Page no.
1	Introduction	1
2	Literature Review	2
3	Research Gap	9
4	Aim	10
5	Objectives	10
6	Research Hypotheses	10
7	Research Methodology	11
8	Statistical Plan	13
9	Expected Outcomes	13
10	References	14

Statement by the Candidate

I wish to state that the work embodied in this synopsis titled “AI based Phenotype and Crop Yield prediction in Climate Smart Agriculture Analytics” forms my own contribution to the work carried out under the guidance of Dr. Amiya Kumar Tripathy at the Shah and Anchor Kutchhi Engineering College. I declare that this written submission represents my ideas in my own words and where others ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.

(Candidate Signature)

Name: Archana Kamlakant Salaskar
PRN Number: 124PHCM1001

1. Introduction

Agriculture plays a crucial role in ensuring global food security, supporting economies, and sustaining the livelihoods of billions. However, the sector is increasingly threatened by climate change, which has brought unpredictable weather patterns, extreme temperatures, and water shortages. To address these challenges, adopting climate-smart agricultural practices is essential for maintaining productivity while reducing environmental impact[34][35]. A promising solution lies in leveraging Artificial Intelligence (AI) to improve crop yield predictions and phenotype analysis across diverse agro-ecosystems.

Worldwide, significant efforts are being made to develop climate-resilient crops and enhance agricultural efficiency. Advanced AI models, integrated with remote sensing and GIS technologies, are transforming how crop health is monitored, yields are predicted, and resources are optimized. Leading nations such as the United States and China are at the forefront of AI-driven agriculture, utilizing multi-modal dataset including genomic, phenotypic, and environmental data to enhance crop performance. Additionally, remote sensing technologies like Synthetic Aperture Radar (SAR) have revolutionized yield forecasting by offering real-time insights into crop growth under diverse climatic conditions. [32][33].

In India, agriculture contributes to 15% of the GDP and employs over 40% of the population. However, the country is highly vulnerable to climate change, with nearly 70 million hectares of crops lost to heatwaves and water stress between 2015 and 2021[27]. Recognizing these challenges, India has introduced over 1,900 climate-resilient crop varieties since 2014, including high-yielding cereals, pulses, oilseeds, and horticultural crops [28]. The Indian Council of Agricultural Research (ICAR) has spearheaded initiatives to develop crops that can withstand droughts, heat stress, and nutrient deficiencies [29]. Despite these advancements, there is a pressing need for scalable AI-driven solutions that integrate multi-source datasets with remote sensing and GIS technologies to enhance crop yield predictions and optimize climate-smart agriculture practices [30].

Table: Few Old and New Climate-Resilient Crop Varieties in India [27][28][31]

Sr No.	Crop Name with Variety and Year	Climate-Resilient Trait
1	Chickpea (Drought-Tolerant, 2021)	Tolerant to drought
2	Pigeonpea (Wilt and Mosaic Resistant, 2021)	Resistant to wilt and sterility mosaic disease
3	Soybean (Early Maturing, 2021)	Early maturing
4	Rice (CR Dhan 810, 2024)	Submergence tolerance, pest resistance
5	Millet (Jowar CSH 30, 2024)	Early maturing, resistant to five diseases

6	Amaranth (Jodhpur Rajgira 2, 2024)	High grain yield, biofortified, withstands temperature fluctuations, low water requirement
7	Rice (Low Phosphorus-Tolerant Varieties, 2024)	Tolerant to phosphorus deficiency, water-saving (30%), lower methane 8emissions
8	Wheat (HI1665 Pusa Gehun Sharbati, 2024)	Heat stress tolerant, suited for dry areas
9	Green Gram (NCR-Specific, 2024)	Heat-tolerant, early maturing
10	Finger Millet (High-Calcium Ragi, 2024)	Biofortified with high calcium content
11	Chickpea (Drought-Tolerant, 2021)	Tolerant to drought
12	Pigeonpea (Wilt and Mosaic Resistant, 2021)	Resistant to wilt and sterility mosaic disease

The integration of AI with advanced technologies such as remote sensing and GIS is set to revolutionize agriculture by facilitating precise resource management, early detection of crop failures, and data-driven decision-making for policymakers. This research supports global food security by developing resilient AI models capable of adapting to diverse agro-climatic conditions while promoting sustainable agricultural practices in India and beyond.

2. Literature Review

Wang et al. [1] to predict county-level winter wheat yields across the Conterminous United States developed machine learning models to by integrating multi-source data, with historical yield data , satellite imagery, climate records, and soil maps. Comparing six machine learning algorithms, they identified AdaBoost as the effective, with an R^2 of 0.86. Their findings emphasized the advantages of multi-source data over single-source approaches, enabling reliable yield predictions ($R^2 > 0.84$) up to 2.5 months before harvest. Here Significance of soil properties underscored and demonstrated that machine learning methods outperformed traditional linear regression models in yield prediction.

Williamson et al. [2] explore the data management challenges associated with applying AI in plant and agricultural research. They emphasize the critical role of FAIR (Findable, Accessible, Interoperable, and Reusable) principles in ensuring efficient data utilization. The study highlights the need for enhanced data infrastructure, standardization, and integration to support AI-driven advancements in agriculture. Additionally, the authors underscore the importance of metadata, data provenance, and ethical considerations in AI applications. The paper concludes with a call for greater collaboration among researchers, institutions, and policymakers to overcome these challenges and fully harness AI's potential in agricultural science.

Zhai et al. [3] investigated to estimate maize chlorophyll content from UAV remote sensing data the use of multisource feature fusion and stacking ensemble learning. They extracted spectral, thermal, and structural features from UAV imagery across three maize growth stages and compared the performance of ridge regression, LightGBM, random forest, and stacking ensemble models. Their findings revealed that integrating multisource features significantly enhanced estimation accuracy, yielding R^2 values between 0.699 and 0.754. Notably, the stacking ensemble model outperformed traditional machine learning approaches, particularly when leveraging multisource data.

Aslan et al. [4] present a comprehensive review of AI techniques for crop yield estimation using Sentinel-2 satellite data. Analyzing 30 research papers from 2019 to 2024, the study highlights the growing adoption of Sentinel-2 imagery in agricultural applications. The authors provide a detailed summary of these studies, covering crop types, vegetation indices, and AI methodologies. In recent years, they observe a shift toward advanced machine learning techniques, including deep learning and ensemble methods. The review underscores the potential of integrating Sentinel-2 data with AI to enhance agricultural sustainability and efficiency while also identifying key limitations and areas for future research.

Lu et al. [5] conducted a study in China utilizing multi-source data and advanced deep learning techniques to predict yields for maize, rice, and soybeans. They employed a CNN-LSTM-Attention model, which outperformed outmoded machine learning methods such as Random Forest and XGBoost. The research integrated various data sources, environmental variables and photosynthetically active parameters including vegetation indices emphasizing the newly introduced kNDVI. The findings demonstrated the CNN-LSTM-Attention model's ability to effectively capture the spatial and temporal complexities of agricultural data, resulting in more accurate yield predictions. Additionally, the study highlighted kNDVI’s potential to enhance yield estimation across different crop types.

Yang et al. [6] analyzed datasets covering central and eastern agricultural regions of China using an in situ monitoring network of 838 stations. They evaluated the accuracy and spatial-temporal variations of soil moisture estimates from satellite products (AMSR2 and ESA-CCI) and reanalysis datasets (ERA5, GLEAM, and GLDAS). The study found that while most products effectively captured soil moisture dynamics, ERA5 performed the best. Overall, reanalysis datasets outperformed satellite-based products, exhibiting stronger spatial-temporal correlations and higher accuracy. Additionally, the authors examined the influence of land use types and wet-dry conditions on product performance, noting lower accuracy in forested, grassland, and arid regions.

Elbasi et al. [7] conducted a comprehensive study on optimizing agricultural data analysis through AI-driven decision-making. They introduced a novel model that utilizes historical agricultural datasets and machine learning algorithms to improve prediction and classification accuracy. Their methodology involved collecting diverse datasets, applying multiple classifiers, and recording accuracy metrics in a CSV file, which was then used to predict the best-performing algorithms for new, unseen data. Experimental results demonstrated significant improvements in predictive accuracy and resource efficiency, with the system effectively identifying near-optimal machine learning models for crop data. The study underscores the potential of this approach in systematically selecting the most suitable algorithms for specific agricultural datasets, enhancing computational efficiency and scalability in smart farming solutions.

T. Yu et al.[8] The study explores the application of Gradient Boosting Decision Trees (GBDT) for predicting phenotypic traits from high-dimensional genomic data in wheat. It compares GBDT with traditional genomic selection (GS) models, including RRBLUP and Bayesian methods, assessing prediction accuracy and computational efficiency. Results indicate that while Bayesian models achieve higher accuracy after extensive iterations, GBDT offers superior computational efficiency and competitive performance for specific traits. The study highlights the impact of marker selection on model performance and suggests that optimizing machine learning techniques can enhance genomic prediction. Future research directions include integrating deep learning and refining feature selection for improved GS accuracy.

Mushtaq et al. [9] review explores the role of artificial intelligence (AI) in advancing for sustainable food security in wheat breeding. It highlights how AI technologies, including machine learning, deep learning, and genomic selection, are transforming wheat breeding by improving efficiency, precision, and sustainability. The study examines AI's applications in high-throughput phenotyping, gene functional analysis, and the rapid identification of stress-tolerant genotypes. AI-driven models are shown to enhance prediction accuracy for crop productivity and disease resistance compared to traditional methods. The authors emphasize that integrating AI into wheat breeding is essential for developing climate-resilient, high-yielding cultivars to meet global food security challenges amid climate change.

C. Egea-Gilabert et al. [10] This special issue focuses on genotype × environment interactions (GEI) in crop breeding, presenting 14 original research articles. The studies explore diverse crops, including cereals, legumes, and vegetables, utilizing approaches such as genomic selection, multi-environment trials, and statistical modeling. Key topics include yield stability, environmental adaptation, and responses to biotic and abiotic stresses. The issue underscores the significance of GEI in developing resilient crop varieties suited to different pedoclimatic

conditions and low-input farming systems. It offers valuable insights for geneticists, breeders, statisticians, and agronomists striving to enhance crop improvement amid climate change and global population growth.

X. Huang et al. [11] This study assessed genotype to environment interactions for yield traits and adaptability in rice cultivars grown under subtropical, temperate, and tropical conditions. Researchers evaluated 89 rice varieties over two years across three locations, analyzing food grain yield and its traits. Results showed that genotype to environment interactions contributed 12.4–19.6% of the variation, while environmental factors accounted for 23.6–69.6%, highlighting their significant influence. For grain yield, genotype to environment interactions explained 37.1% of the variation, underscoring strong environmental effects. Using WAASBY indexing, the study identified 19 high-yielding, stable genotypes suitable for diverse climates, offering valuable insights for breeding programs focused on improving yield potential and adaptability.

Liersch et al. [12] This study evaluated 25 winter oilseed rape genotypes to analyze genotype × environment (G×E) interactions for six seed traits across six environments in Poland. The additive main effects and multiplicative interaction (AMMI) model, researchers examined oil content, fatty acid composition, glucosinolate content, and protein content. Results revealed significant G×E interactions for altogether traits, with the environment strongly influencing oil and protein content, while genotype had the greatest impact on fatty acid composition and glucosinolate content. The AMMI analysis identified four highly stable genotypes with optimal trait values, making them promising candidates for future breeding programs.

Hilmarsson et al. [13] This study examined genotype × environment (G×E) interactions for spring barley yield and thousand-kernel weight (TKW) in Iceland using AMMI, factorial regression, and linear mixed models. Findings showed that soil fertility was the key factor influencing G×E for yield, with 6-row barley performing better in high-fertility soils than 2-row barley. For TKW, daily mean temperature played a major role, with distinct responses observed between the two barley types along the temperature gradient. Based on yield performance, for barley breeding the authors recommended Iceland as a single mega-environment but noted that distinct mega-environments could be defined if TKW became the primary breeding focus.

Zhenyu Ma et al.[14], in crop monitoring this review provides an in-depth analysis of sensing technologies for high-throughput plant phenotyping. It explores key phenotyping parameters, communication procedures, and data management systems used in the field. Various sensor types, including visible spectrum, infrared, multi-spectral, hyperspectral, depth, LiDAR, and fluorescence sensors, are discussed along with their applications and limitations. The paper also

examines ground-based and aerial phenotyping platforms, highlighting their benefits and challenges. Concluding with an overview of current limitations, the review emphasizes the need for integrated systems that seamlessly combine plant measurement, data transmission, management, and visualization for improved efficiency in crop monitoring.

Sah et al. [15] study was conducted to predict rice yields using machine learning models that integrated biophysical constraints with SAR as well as optical remote sensing data. The research, carried out across two rice crop seasons in Udham Singh Nagar, Uttarakhand, India, focused on yield predictions at 45, 60, and 90 days after transplanting (DAT). Nine machine learning models were evaluated, including Elastic Net, Random Forest, and eXtreme Gradient Boosting (XGB). For summer rice, XGB consistently outperformed other models, while the best-performing models for kharif rice varied across growth stages. Results showed that prediction accuracy improved as harvest approached, with the most reliable yield estimates at 90 DAT.

Joshi et al. [16] This study explored the use of artificial intelligence (AI) and satellite-based remote sensing for soybean yield prediction. Researchers compared various AI models, including deep neural networks, support vector machines, random forests, LASSO, and AdaBoost, using Planet Scope satellite imagery captured at different growth stages. Results showed that deep neural networks models outperformed other approaches, with prediction accuracy improving as more growth stage data were incorporated. Satellite images from later reproductive stages (R4/R5) were found to be the most crucial for yield predictions. While the study highlighted the potential of integrating AI and remote sensing for field-scale soybean yield estimation, model transferability across years and locations remained a challenge.

Wang et al. [17] this review examines the use of remote sensing (RS) technology with machine learning (ML) algorithms in precision agriculture (PA). It highlights hyperspectral RS as the most widely used method, comprising over 30% of applications, while UAV technology shows significant potential at around 24%. Among ML algorithms, support vector machines (SVM) dominate with over 20% usage, followed by random forests at 18%. The study explores applications in agricultural monitoring, disease detection, land management, and crop yield prediction. Additionally, it discusses key challenges, such as data acquisition, model interpretability, and generalization, while emphasizing future trends toward intelligence, automation, and sustainable agricultural development.

Abdelraouf M. Ali et al. [18] review provides an in-depth crop yield prediction analysis using multi-sensor remote sensing techniques. It explores various remote sensing tools, including satellite imagery, aerial photography, RADAR, LIDAR, and field sensors, and their applications in agricultural monitoring and yield estimation. The study highlights the benefits of high-

resolution satellite data, cost-effective UAVs, and field hyperspectral measurements in advancing agricultural remote sensing. Additionally, it examines vegetation indices and their role in estimating crop biophysical parameters. The authors conclude that integrating satellite data with ground-based hyperspectral measurements is essential for achieving more accurate and comprehensive crop yield predictions.

Abhasa Joshi et al. [19], This systematic review examined 90 studies on deep learning (DL) and remote sensing (RS) applications in crop mapping and yield prediction. Platforms, sensors, input features, training data, study site distribution, output scales, architectures, frameworks, evaluation metrics, and model performance are analyzed as key. The review found that satellite-based sensors were the most utilized RS data sources, with MODIS commonly used for yield prediction and Sentinel-1, Sentinel-2, and Landsat for crop mapping. Convolutional Neural Networks (CNNs) emerged as the most popular DL architecture. Identified challenges included limited training data, the need for generalizable models, and transparency in predictions. The authors recommended future research on improving data availability, developing efficient and interpretable models, and exploring unsupervised learning techniques.

I. Malashin et al.[20], study focuses on predicting sustainable crop yields using deep learning and explainable AI. A deep neural network (DNN) is optimized using genetic algorithms (GAs) to improve accuracy, achieving an R^2 score of 0.92, meaning it explains 92% of yield variation. Explainable AI techniques, such as local interpretable model-agnostic explanations (LIME), help make the model more transparent by showing which features are most important. The study also develops a classification model using a GA-optimized DNN to identify the best crops based on soil and climate data. The results confirm that this approach effectively predicts crop yields and helps select suitable crops for sustainable farming.

Rebez et al. [21], review explores the role of artificial intelligence (AI) in managing heat stress in ruminant livestock. It examines how to monitor and assess heat stress by analyzing animal behavior, physiology, and health using AI technologies, like machine learning, neural networks, and deep learning, are used. The study highlights AI's potential in enhancing heat stress prediction, developing early warning systems, and supporting data-driven decision-making in livestock management. Additionally, it discusses AI applications in evaluating production parameters and the genetic potential of heat-stressed animals. The authors conclude that AI-powered tools present innovative solutions for mitigating heat stress and promoting sustainable livestock production amid climate change.

Raza et al. [22], in this paper the transformative impact of artificial intelligence (AI) on climate-smart agriculture examined, highlighting recent advancements and future prospects. It explores

how AI technologies, like machine learning and deep learning, are revolutionizing agricultural analysis and enhancing climate-resilient farming practices. Key applications include predictive modeling for climate events, optimizing irrigation schedules, and early detection of crop stress and disease. The integration of AI-powered precision agriculture tools, such as autonomous drones and sensor networks, facilitates real-time monitoring and data-driven decision-making. Additionally, the study underscores AI's role in improving post-harvest management, storage, and distribution, ultimately boosting agricultural efficiency and resilience against climate challenges.

Assimakopoulos et al.[23], this review explains how artificial intelligence (AI) helps improve agriculture by boosting productivity, sustainability, and resource efficiency. It discusses how AI technologies are used for land planning, crop selection, resource management, disease detection, yield prediction, and market access. The study also highlights challenges such as limited data access and infrastructure issues. It explores AI's potential to solve global agricultural problems and offers useful insights for farmers, businesses, policymakers, and researchers. The goal is to promote AI adoption and encourage discussions on making agriculture more sustainable.

Dhal et al. [24], explores how AI-driven forecasting models are transforming global food security. It examines how machine learning, deep learning, and time-series forecasting improve farming and food supply chains. The study highlights AI applications in traditional and modern farming, such as hydroponics, aquaponics, and vertical farming. It also discusses how IoT, remote sensing, and blockchain help monitor crops, manage resources, and analyze markets in real time. While showcasing AI's benefits, the paper also addresses challenges like data quality, model scalability, and prediction accuracy, especially in resource-limited areas.

Khan et al. [25], This review examines the role of artificial intelligence (AI) in advancing climate-resilient smart-crop breeding. It explores how AI technologies address challenges in high-throughput phenotyping, gene functional analysis, and data processing for both phenotyping and genotyping. The study highlights AI's potential in integrating genomics, phenomics, and Environics data to enhance breeding strategies for climate-resilient crops. Additionally, it discusses AI's ability to bridge the gap between genotypes and phenotypes, a key challenge in modern breeding. The review underscores AI's impact on accelerating crop improvement programs by enabling rapid gene identification and improving the accuracy of high-throughput phenotyping, genotyping, and envirotyping data.

Rashid et al.[26], review provides an in-depth analysis of crop yield prediction using machine learning, with a particular emphasis on palm oil yield forecasting. It examines a wide range of features and algorithms used in crop yield modeling, evaluating their strengths and limitations.

The study highlights the absence of standardized feature sets and optimal algorithms for palm oil yield prediction. Additionally, it explores machine learning applications in the palm oil industry, including disease detection, growth monitoring, and plantation mapping. The authors propose an integrated predictive framework that combines multiple data sources and agronomic factors. Overall, the review synthesizes existing research and outlines future directions to enhance machine learning-based crop and palm oil yield prediction models.

3. Research Gaps

- Despite advancements, there is a lack of standardized frameworks for integrating diverse datasets such as genomic, phenotypic, environmental, and remote sensing data. Many studies focus on single-source data, limiting the potential to uncover complex relationships between genotype-environment interactions and crop traits.

- Current AI models struggle to fully capture the complexity of genotype-environment interactions due to the high dimensionality of genomic data and the variability in environmental conditions. There is a need for robust multi-modal AI models that can simultaneously process and analyze these interactions across diverse agro-ecosystems.

- While remote sensing and GIS technologies have improved crop yield predictions, challenges remain in achieving high-resolution spatiotemporal predictions. Issues such as limited temporal resolution, and insufficient integration with ground-truth data hinder the accuracy and applicability of these models.

- Existing studies often fail to optimize climate-smart agriculture factors across different agro-ecosystems. There is a gap in developing AI models that can adapt to local conditions while providing actionable insights for farmers under varying climatic scenarios.

4. Aim

To develop an AI-driven robust model that will integrate multi-source datasets, captures genotype-environment interactions, and incorporates Remote Sensing data to optimize climate-smart agriculture practices and predict crop traits as well as yields in a diverse agro-ecosystems.

5. Objectives

1. **To Explore** the multi-source datasets for understanding their distribution, relationships towards the selection and preprocessing for AI-based analysis.
2. **To Design & Develop** an AI model that will capture genotype-environment interactions and uses multi-modal data to predict crop traits.
3. **To Integrate** Remote Sensing and GIS data into the AI model to enhance spatiotemporal predictions for crop yield.
4. **To Optimize & Evaluate** climate-smart agriculture factors that influence the crop performance and robustness of the AI model across different agro-ecosystems.

6. Research Hypotheses (RH)

- RH1: The integration of diverse multi-source datasets (genomic, phenotypic, environmental, Remote Sensing data, etc.) will reveal complex relationships and distributions that, when properly preprocessed, will significantly enhance the accuracy and robustness of AI-based crop trait and yield prediction models.

- RH2: A deep learning model incorporating both convolutional and recurrent neural network, trained on multi-modal data, will more accurately capture genotype-environment interactions and predict crop traits compared to traditional statistical methods or single-modality AI models.

- RH3: The integration of high-resolution Remote Sensing and GIS data into a deep learning model will significantly improve the spatiotemporal accuracy of crop yield predictions across diverse agricultural landscapes, outperforming models that rely solely on ground-based or lower-resolution data sources.

- RH4: Optimize the most influential climate-smart agriculture factors for crop performance across different agro-ecosystems towards adaptability, scalability and applicability of prediction models.

7. Research Methodology

The research methodology for the proposed approach focuses on developing an innovative Model:

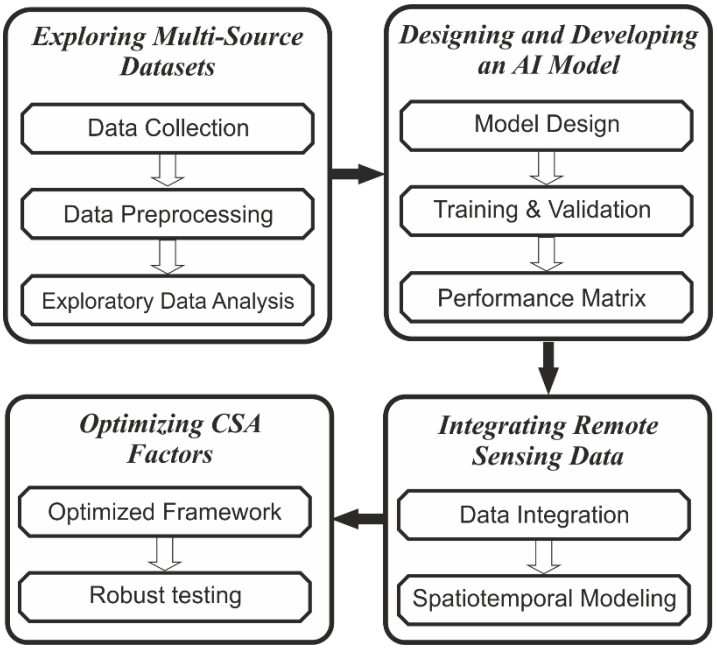
1. Exploring Multi-Source Datasets
   - Data Collection:
     - Gather diverse datasets, including:
       - Genomic data (e.g., SNP markers).
       - Phenotypic data (e.g., yield traits, growth parameters).
       - Environmental data (e.g., soil properties, precipitation, temperature).
       - Remote sensing data (e.g., NDVI).
       - GIS data (e.g., land use maps).
     - Use publicly available sources like TERRA-REF, Genomes to Fields (G2F), and ICAR databases.
   - Data Preprocessing:
     - Handle missing values, outliers, and inconsistencies using imputation techniques.
     - Normalize and standardize datasets to ensure compatibility across different data types.
     - Extract key features such as vegetation indices from remote sensing data.
     - Align spatial and temporal resolutions of datasets using GIS tools.
   - Exploratory Data Analysis:
     - Analyze the distribution of multi-source datasets.
     - Visualize relationships and correlations between variables using statistical tools.

2. Designing and Developing an AI Model
   - Model Design:
     - Develop a multi-modal AI model that integrates:
       - CNNs for remote sensing imagery analysis.
       - RNNs for sequential environmental data processing.
   - Training and Validation:
     - Split datasets into training, validation, and test sets.
     - Train the model using multi-modal inputs (genomic, phenotypic, environmental, and remote sensing data).
     - Use cross-validation techniques to avoid overfitting.
   - Performance Metrics:
     - Evaluate model performance using metrics .
3. Integrating Remote Sensing and GIS Data
   - Data Integration:
     - Combine high-resolution remote sensing imagery with GIS layers for spatial analysis.
     - Extract vegetation indices like NDVI to enhance crop yield prediction accuracy.
   - Spatiotemporal Modeling:
     - Develop spatiotemporal models capable of generating dynamic crop yield predictions at fine spatial resolutions.
     - Incorporate seasonal climate data to account for temporal variations.

4. Optimizing Climate-Smart Agriculture Factors

- Optimization Framework:
  - Identify key climate-smart agriculture factors and optimize it.
- Robustness Testing:
  - Test the AI model across diverse agro-ecosystems with varying climatic conditions.
  - Evaluate model adaptability using unseen datasets from different regions.

Figure 1: Phases for Proposed Research Methodology



8. Statistical Plan

i) Metrics for Evaluation:

- Accuracy, precision, recall, F1-score, rank-1 accuracy

ii) Hypothesis Testing:

- Different tests will be performed to compare the performance of the proposed approach with existing methods.

9. Expected Outcomes:

- A comprehensive understanding of the distribution and relationships within multi-source datasets will be achieved. This will result in the development of a standardized preprocessing pipeline to clean, normalize, and integrate heterogeneous datasets for compatibility in AI-based analysis.

- A robust AI model will be developed that accurately predicts crop traits as well as yield with multi source data integration. This model will demonstrate improved prediction accuracy compared to traditional statistical or single-modal AI approaches.

- The integration of high-resolution remote sensing imagery (e.g., NDVI) with GIS layers (e.g., soil maps, field boundaries) into the AI model will enable enhanced spatiotemporal resolution in crop yield predictions. This will result in dynamic yield maps that account for spatial variability across agro-ecosystems.

- Key climate-smart agriculture factors will be optimized. The resulting robust AI model will maintain high predictive accuracy across diverse agro-ecosystems under varying climatic conditions, providing actionable insights for sustainable farming practices.

10. References

[1] Y. Wang, Z. Zhang, L. Feng, Q. Du, and T. Runge, "Combining Multi-Source Data and Machine Learning Approaches to Predict Winter Wheat Yield in the Conterminous United States," *Remote Sensing*, vol. 12, no. 8, p. 1232, Apr. 2020, doi: https://doi.org/10.3390/rs12081232.

[2] Williamson HF, Brettschneider J, Caccamo M, Davey RP, Goble C, Kersey PJ, May S, Morris RJ, Ostler R, Pridmore T, Rawlings C, Studholme D, Tsaftaris SA, Leonelli S. Data management challenges for artificial intelligence in plant and agricultural research. F1000Res. 2021 Apr 27;10:324. doi: 10.12688/f1000research.52204.2. PMID: 36873457; PMCID: PMC9975417.

[3] W. Zhai, C. Li, Q. Cheng, F. Ding, and Z. Chen, "Exploring Multisource Feature Fusion and Stacking Ensemble Learning for Accurate Estimation of Maize Chlorophyll Content Using Unmanned Aerial Vehicle Remote Sensing," *Remote Sensing*, vol. 15, no. 13, pp. 3454–3454, Jul. 2023, doi: https://doi.org/10.3390/rs15133454.

[4] Muhammet Fatih Aslan, Kadir Sabanci, and B. Aslan, "Artificial Intelligence Techniques in Crop Yield Estimation Based on Sentinel-2 Data: A Comprehensive Survey," *Sustainability*, vol. 16, no. 18, pp. 8277–8277, Sep. 2024, doi: https://doi.org/10.3390/su16188277

[5] Lu, J.; Li, J.; Fu, H.; Tang, X.; Liu, Z.; Chen, H.; Sun, Y.; Ning, X., "Deep Learning for Multi-Source Data-Driven Crop Yield Prediction in Northeast China," *Agriculture*, vol. 14, no. 6, pp. 794–794, May 2024, doi: https://doi.org/10.3390/agriculture14060794.

[6] Yang, Y.; Zhang, J.; Bao, Z.; Ao, T.; Wang, G.; Wu, H.; Wang, J. "Evaluation of Multi-Source Soil Moisture Datasets over Central and Eastern Agricultural Area of China Using In Situ Monitoring Network," *Remote Sensing*, vol. 13, no. 6, pp. 1175–1175, Mar. 2021, doi: https://doi.org/10.3390/rs13061175.

[7] Ersin Elbasi, N. Mostafa, C. Zaki, Zakwan AlArnaout, A. E. Topcu, and L. Saker, "Optimizing Agricultural Data Analysis Techniques through AI-Powered Decision-Making Processes," *Applied Sciences*, vol. 14, no. 17, pp. 8018–8018, Sep. 2024, doi: https://doi.org/10.3390/app14178018.

[8] T. Yu *et al.*, "Predicting phenotypes from high-dimensional genomes using gradient boosting decision trees," *IEEE Access*, pp. 1–1, 2022, doi: https://doi.org/10.1109/access.2022.3171341.

[9] Muhammad Ahtasham Mushtaq, Ahmed, and Y. Zeng, "Applications of Artificial Intelligence in Wheat Breeding for Sustainable Food Security," *Sustainability*, vol. 16, no. 13, pp. 5688–5688, Jul. 2024, doi: https://doi.org/10.3390/su16135688.

[10] C. Egea-Gilabert, M. A. Pagnotta, and P. Tripodi, "Genotype × Environment Interactions in Crop Breeding," *Agronomy*, vol. 11, no. 8, p. 1644, Aug. 2021, doi: https://doi.org/10.3390/agronomy11081644.

[11] X. Huang, S. Jang, B. Kim, Z. Piao, E. Redona, and H.-J. Koh, "Evaluating Genotype × Environment Interactions of Yield Traits and Adaptability in Rice Cultivars Grown under Temperate, Subtropical and Tropical Environments," *Agriculture*, vol. 11, no. 6, p. 558, Jun. 2021, doi: https://doi.org/10.3390/agriculture11060558.

[12] Liersch, A.; Bocianowski, J.; Nowosad, K.; Mikołajczyk, K.; Spasibionek, S.; Wielebski, F.; Matuszczak, M.; Szała, L.; Cegielska-Taras, T.; Sosnowska, K.; et al., "Effect of Genotype × Environment Interaction for Seed Traits in Winter Oilseed Rape (Brassica napus L.)", *Agriculture*, vol. 10, no. 12, pp. 607–607, Dec. 2020, doi: https://doi.org/10.3390/agriculture10120607.

[13] Yue, H.; Gauch, H.G.; Wei, J.; Xie, J.; Chen, S.; Peng, H.; Bu, J.; Jiang, X. , "Genotype

by Environment Interaction Analysis for Grain Yield and Yield Components of Summer Maize Hybrids across the Huanghuaihai Region in China," *Agriculture*, vol. 12, no. 5, pp. 602–602, Apr. 2022, doi: https://doi.org/10.3390/agriculture12050602.

[14] Ma, Zhenyu & Rayhana, Rakiba & Feng, Ke & Liu, Zheng & Xiao, George & Ruan, Yuefeng & Sangha, Jatinder, "A Review on Sensing Technologies for High-Throughput Plant Phenotyping," *IEEE Open Journal of Instrumentation and Measurement*, vol. 1, pp. 1–21, 2022, doi: https://doi.org/10.1109/ojim.2022.3178468.

[15] S. Sah, D. Haldar, R. N. Singh, B. Das, and A. S. Nain, "Rice yield prediction through integration of biophysical parameters with SAR and optical remote sensing data using machine learning models," *Scientific Reports*, vol. 14, no. 1, Sep. 2024, doi: https://doi.org/10.1038/s41598-024-72624-4

[16] D. R. Joshi, Sharon A. Clay, Prakriti Sharma, Hossein Moradi Rekabdarkolaee, Tulsi Kharel, Donna M. Rizzo, Resham Thapa, David E. Clay, "Artificial Intelligence and Satellite Based Remote Sensing can be used to Predict Soybean (Glycine max) Yield," *Agronomy Journal*, Sep. 2023, doi: https://doi.org/10.1002/agj2.21473.

[17] J. Wang, Y. Wang, G. Li, and Z. Qi, "Integration of Remote Sensing and Machine Learning for Precision Agriculture: A Comprehensive Perspective on Applications," *Agronomy*, vol. 14, no. 9, p. 1975, Sep. 2024, doi: https://doi.org/10.3390/agronomy14091975.

[18] Abdelraouf M. Ali, Mohamed Abouelghar, A.A. Belal, Nasser Saleh, Mona Yones, Adel I. Selim, Mohamed E.S. Amin, Amany Elwesemy, Dmitry E. Kucher, Schubert Maginan, Igor Savin, "Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article)," *The Egyptian Journal of Remote Sensing and Space Science*, vol. 25, no. 3, pp. 711–716, Dec. 2022, doi: https://doi.org/10.1016/j.ejrs.2022.04.006.

[19] A. Joshi, B. Pradhan, S. Gite, and S. Chakraborty, "Remote-Sensing Data and Deep-Learning Techniques in Crop Mapping and Yield Prediction: A Systematic Review," *Remote Sensing*, vol. 15, no. 8, p. 2014, Jan. 2023, doi: https://doi.org/10.3390/rs15082014.

[20] I. Malashin, Vadim Tynchenko, Andrei Gantimurov, Vladimir Nelyub, Aleksei Borodulin, and Yadviga Tynchenko, "Predicting Sustainable Crop Yields: Deep Learning and Explainable AI Tools," *Sustainability*, vol. 16, no. 21, pp. 9437–9437, Oct. 2024, doi: https://doi.org/10.3390/su16219437.

[21] Rebez, E.B.; Sejian, V.; Silpa, M.V.; Kalaignazhal, G.; Thirunavukkarasu, D.; Devaraj, C.; Nikhil, K.T.; Ninan, J.; Sahoo, A.; Lacetera, N.; *et al.*, "Applications of Artificial Intelligence for Heat Stress Management in Ruminant Livestock," *Sensors*, vol. 24, no. 18, pp. 5890–5890, Sep. 2024, doi: https://doi.org/10.3390/s24185890.

[22] Raza, A.; Shahid, M.; Safdar, M.; Zaman, M.; Sabir, R. "The Role of Artificial Intelligence in Climate-Smart Agriculture: A Review of Recent Advances and Future Directions", in Proceedings of the 2nd International Online Conference on Agriculture, 1–15 November 2023, MDPI: Basel, Switzerland, doi:10.3390/IOCAG2023-16877

[23] F. Assimakopoulos, C. Vassilakis, D. Margaris, K. Kotis, and D. Spiliotopoulos, "Artificial Intelligence Tools for the Agriculture Value Chain: Status and Prospects," *Electronics*, vol. 13, no. 22, p. 4362, Nov. 2024, doi: https://doi.org/10.3390/electronics13224362.

[24] S. B. Dhal and D. Kar, "Transforming Agricultural Productivity with AI-Driven Forecasting: Innovations in Food Security and Supply Chain Optimization," *Forecasting*, vol. 6, no. 4, pp. 925–951, Oct. 2024, doi: https://doi.org/10.3390/forecast6040046.

[25] Khan, M.H.U.; Wang, S.; Wang, J.; Ahmar, S.; Saeed, S.; Khan, S.U.; Xu, X.; Chen, H.; Bhat, J.A.; Feng, X., M. H. U. Khan *et al.*, "Applications of Artificial Intelligence

in Climate-Resilient Smart-Crop Breeding,” *International Journal of Molecular Sciences*, vol. 23, no. 19, p. 11156, Sep. 2022, doi: https://doi.org/10.3390/ijms231911156.

[26] M. Rashid, B. S. Bari, Y. Yusup, M. A. Kamaruddin and N. Khan, "A Comprehensive Review of Crop Yield Prediction Using Machine Learning Approaches With Special Emphasis on Palm Oil Yield Prediction," in *IEEE Access*, vol. 9, pp. 63406-63439, 2021, doi: 10.1109/ACCESS.2021.3075159.

[27] A. Chaturvedi, hindustantimes.com, “PM Modi releases climate resilient crop varieties: How will they help India,” *Hindustan Times*, Sep. 28, 2021. https://www.hindustantimes.com/india-news/pm-modi-releases-climate-resilient-crop-varieties-how-will-they-help-india-101632817619148.html. (accessed Feb. 12, 2025).

[28] V. Kurtkoti, "India adapts to climate change with new resilient crops: Cereals and oilseeds among 109 varieties given to farmers," *News*, Sep. 6, 2024. [Online]. Available: [https://www.nature.com/articles/d44151-024-00143-2].

[29] V. Kurtkoti, "India looks to sow more climate-resilient crops as it braces for more heatwaves, water crises," *Channel News Asia*, Aug. 27, 2024. [Online].Available: https://www.channelnewsasia.com/asia/india-sow-climate-resilient-crops-rice-wheat-farmers-4759371?utm_source=perplexity.

[30] L. Navi, K. Kushal, G. K. Gangadhar, and S. Arangaji, "Yield prediction in field crops through satellite and remote sensing," *The Pharma Innovation Journal*, vol. 11, no. 12S, pp. 2437–2440, 2023

[31] Indian Council of Agricultural Research, "PM Dedicates 109 Climate-Resilient and Bio-Fortified Varieties of Crops," ICAR, 12 Jan 2025. [Online]. Available: https://icar.gov.in/pm-dedicates-109-climate-resilient-and-bio-fortified-varieties-crops.

[32] Cembrowska-Lech, D.; Krzemińska, A.; Miller, T.; Nowakowska, A.; Adamski, C.; Radaczyńska, M.; Mikiciuk, G.; Mikiciuk, M. An Integrated Multi-Omics and Artificial Intelligence Framework for Advance Plant Phenotyping in Horticulture. *Biology* **2023**, *12*, 1298. https://doi.org/10.3390/biology12101298

[33] S. Iniyan and R. Jebakumar, “Mutual Information Feature Selection (MIFS) Based Crop Yield Prediction on Corn and Soybean Crops Using Multilayer Stacked Ensemble Regression (MSER),” *Wireless Personal Communications*, Jun. 2021, doi: https://doi.org/10.1007/s11277-021-08712-9.

[34] D. K. Gupta, A. Pagani, P. Zamboni, and A. K. Singh, “AI-powered revolution in plant sciences: advancements, applications, and challenges for sustainable agriculture and food security,” vol. 2, no. 5, pp. 443–459, Aug. 2024, doi: https://doi.org/10.37349/eff.2024.00045.

[35] SS. Mahdi and J. Ali, “Harnessing the power of machine learning for crop improvement and sustainable production,” *Frontiers in Plant Science*, vol. 15, Aug. 2024, doi: https://doi.org/10.3389/fpls.2024.1417912.